DOCUMENT RESUME

ED 345 228 CS 213 244

AUTHOR Gould, Christopher
TITLE Assessment in a Social Context: Grading as an Interpretive Community.
PUB DATE Mar 92
NOTE 13p.; Paper presented at the Annual Meeting of the Conference on College Composition and Communication (43rd, Cincinnati, OH, March 19-21, 1992).
PUB TYPE Reports - Research/Technical (143) -- Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College English; *Freshman Composition; *Grading; Higher Education; *Reader Response; *Writing Evaluation; Writing Instruction; Writing Research
IDENTIFIERS Alternative Assessment; *Interpretive Communities

ABSTRACT

Three teachers of first-year composition used cross-grading as a way of extending the student's grasp of interpretive communities as arbiters of value as well as creators of meaning. Students in six sections (two experimental groups) approached the English 101 Common Final in the same manner, discussing a published article and sharing their preliminary writing before completing a final draft during the examination period. In a practice run, students in Group B observed the three teachers sharing freewritten responses to a published article as a preliminary to composing a polished essay. Both groups saw the teachers' freewrites and polished essays, but only Group B witnessed the verbal negotiations of this "interpretive community." Results showed that: (1) students in Group B did not write better essays on the Common Final than those in Group A; (2) students in Group B may have developed a better understanding of reading and interpretative communities; (3) teachers probably graded student essays more fairly and consistently as a result of having constituted themselves as an interpretive community in front of classes, reaching a rough consistency in grading about 90% of the time; and (4) students in Group B, as evidenced both in their journals and in their quantitative course evaluations, felt better about grading procedures than those in Group A. Evaluation can be demystified when cross-grading partners define themselves as an interpretive community. Cross-graders can demonstrate their reading strategies and acknowledge their critical biases, thus entering into a dialogue that enriches both students and teachers. (SR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Christopher Gould

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Christopher Gould
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED345228

Assessment in a Social Context: Grading as an Interpretive Community

The social dimension of literacy is now a commonplace in composition study, and one practical result is the implementation of reader-response theory in English 101. Pedagogical works often refer to the composition class as an interpretive community, and many instructors invite first-year students to interrogate the conventions of academic discourse.

Welcome as these developments are given the decline of cognitive-process theory, impediments remain. One is the persistence of institutional requirements that militate against active reading. Two of the "core requirements" of English 101 in our department, for example, are "a series of on-going exercises in summary and paraphrase" and a final examination that asks students to read and respond to a published article, beginning with "a one- or two-sentence summary of the article's main idea." Both these assignments, unless deliberately adapted by the instructor--adapted in such a way as to subvert the impulse behind their inclusion in the departmental syllabus--reinforce the notion that meaning resides within the text.

Another impediment is the contradiction between the instructor's role as evaluator of student work and her efforts to enfranchise the class as a community of readers, writers, and critical thinkers. This particular dilemma was underscored recently when one of our more popular and theoretically informed colleagues received the following comment in a student evaluation: "The teacher's grading standards are different from those of the class." What once might have been dismissed as a

CS213244

BEST COPY AVAILABLE

petulant complaint becomes an ingenuous observation, if not a compelling revelation.

The study we are about to describe addresses these concerns. First, we wanted to find a way to engage first-year students in the kind of reading implied by Stanley Fish's admonition that "there is not a single way of reading that is correct or natural, only 'ways of reading' that are extensions of community perspectives" (16). We hoped to model this kind of reading by constituting ourselves as an interpretive community of three educated adults and revealing the negotiations through which that community might arrive at some notion of what a particular text "means." Second, we hoped to complicate the usual views of assessment--either as a disinterested application of objective criteria or as the arbitrary exercise of idiosyncratic notions of "good writing." Accordingly, we used cross-grading as a way of extending our students' grasp of interpretive communities as arbiters of value as well as creators of meaning. In other words, by allowing students to observe how the three of us constructed meaning collaboratively, we hoped to elucidate--for ourselves as well as for them--how we might evaluate the interpretive essays they would write in response to the English 101 Common Final. Doing this, we surrendered the reassuring myths of holistic scoring--strict objectivity, inter-rater reliability, perhaps even valid assessment--all of which appear problematic within the context of reader-response theory. In short, we trusted the force of that theory, our resolve to enact it without compromise, and our students' good will.

A description of the English 101 Common Final is in order. During the next-to-last class meeting of the term, students in all fifty-some sections of the course receive a thousand-word published article to be read and discussed in class. (Newsweek's "My Turn" column is a frequently used source.) The Common Final "prompt," distributed a week earlier in conjunction with a practice run, instructs students to write an essay in which they either "agree or disagree with the writer's controlling idea (thesis or main idea)." Essays are to be developed with "examples, illustrations, and reasons from your own experience, reading, and thinking" and are to open with "a one- or two-sentence summary of the article's main idea." The last regular class meeting is set aside for invention and preliminary composing, leaving a three-hour examination period for polishing a 500-word final draft.

Since each of us taught two sections of approximately thirty students, we created two separate groups for purposes of comparison. (The two groups proved remarkably even in terms of semester-grade average: 2.63 vs. 2.68 on a four-point scale.) Both groups approached the Common Final in the conventional manner, discussing the published article and sharing their preliminary writing in peer groups before completing a final draft during the examination period. Both groups also did a practice run, in which we ourselves participated, using an article chosen for a previous semester. The difference between the two groups lay in our handling of this practice run. Students in Group B observed the three of us sharing freewritten

responses to the published article as a preliminary to composing a polished essay that addressed the Common Final prompt. Both groups saw our freewrites and our polished essays, but only Group B witnessed the verbal negotiations of our "interpretive community"--negotiations that subtly altered our interpretation of the published article and thus influenced our final drafts.

The results of our inquiry are best summarized in relation to four fairly simple questions.

First, the inevitable one: Did students in Group B write better essays than students in Group A? No. Remarkably, in fact, grade averages for the Common Final were identical: 2.67 (virtually the same as the semester averages cited above). Any disappointment over this finding, however, should be mitigated by Knoblauch and Brannon's critique of the "myths about evaluation and improvement" (151-71).

A more provocative question is whether students in Group B developed any better understanding of reading and interpretative communities. The answer, based on our reading of their journals, is a hesitant maybe. Most responded favorably to what we did, and several said exactly what we hoped to hear; for example:

> I found this approach to the assignment valuable because it's helped me see how I can read an article, really think about it, and form my own opinion rather than have someone tell me "This is what you should have seen."
>
> Each of the three readers showed how to relate personal experience to the text and make it work.

> I feel that they [the three instructors] are trying to help you write better on a topic you may not know much about by getting you to think and relate that topic to other things you do know a lot about.
>
> It's amazing to see how many different ideas can come from the same article. It goes to show you how differently readers can view the same text. I think each of the three learned something new after hearing the responses of the others.

Gratifying as such responses were, we learned just as much from the more equivocal assessments. Several students, while applauding our methods, acknowledged confusion. One student confessed:

> Some of the views they took, well, I just couldn't see how they were derived from the text. For example, when Ms. Callahan and Mr. Gould were talking about their kids, I just couldn't see the connection. I guess it shows how broad their views are and how narrow mine are.

This student's last sentence identifies a problem anticipated but not fully overcome: a few students perceived our interpretive negotiations as an unattainable ideal--the way "intellectuals" read. In one or two cases, this concern was tied to grades and examsmanship--the fear that we might, in the words of one student, "expect us to do work like any of you three." For another student, however, the concern was more disinterested: "I found myself feeling like I was watching one of those public

television shows where intellectuals discuss a topic. I tried to tune in as much as possible, but it was hard."

Predictably, about five students openly resisted our methods, preferring a more directive, formalist approach; and one rejected the whole concept of collaboration on the grounds that "freewriting is extremely too personal to discuss with anyone else."

In response to the question of whether we made students better readers, we can claim only to have introduced the possibility--to have initiated a process reinforced, we fear, in few classes outside the English department.

Evidence that our undertaking may have been worthwhile rests more on its implications regarding assessment. A third question to be addressed, then, is whether we graded student essays more fairly and consistently after having constituted ourselves as an interpretive community in front of three of our classes. In this case, the answer is an emphatic probably.

Before supporting that guarded assertion, we must describe our cross-grading procedure. Each of us brought to the session 57 or 58 essays, each of which was to be graded holistically (A to F, no pluses or minuses) by both the other instructors. The first reader's grade was concealed from the second reader, and our only attempt to calibrate scoring was a brief exchange of "range finders" during the opening minutes of the session. After three and a half hours, all 173 essays had been scored twice, which meant that we averaged less than two minutes per reading.

Scores coincided 116 times (67%) and diverged by one letter

grade in 51 instances (30%). (Scores for only six essays diverged by two letter grades and none by more than two.) Granting that many of the 51 near-agreements were truly borderline cases, we reached a rough kind of consistency perhaps 90% of the time--a respectable figure given Cooper's claim of "scoring reliabilities in the high eighties and low nineties" after fairly elaborate calibration techniques (18). Noteworthy, too, is Cooper's explanation for lapses in reliability, for which he quotes Follman and Anderson:

> It may now be suggested that the unreliability usually obtained in the evaluation of essays occurs primarily because raters are to a considerable degree heterogeneous in academic background and have had different experiential backgrounds which are likely to produce different attitudes and values which operate significantly in their evaluations of essays. The function of a theme evaluation procedure, then, becomes that of a sensitizer or organizer of the rater's perception and gives direction to his attitudes and values. (19)

We wish to suggest that our negotiations as an interpretive community brought into the open some of the experiential differences to which Follman and Anderson refer, thus minimizing their distorting effects.

Of course reliability, as Cooper uses the term, involves a great deal more than getting two readers to put the same score on a piece of writing--issues such as whether or not an agreed upon

score is indicative of anything. In the present instance, those issues revolve around the writing task set by the English 101 Common Final and the circumstances under which it is administered--matters about which we have already expressed some reservation.

We do, however, wish to introduce one final concern, addressed in a fourth question: Did the students in Group B feel better, on the whole, about our grading procedures than did the students in Group A? In other words, although the essays of all students were cross-graded in the same manner, did the students in Group B better understand and appreciate the procedure for having observed us engaged as an interpretive community? Reading their journals, we felt that they did; however, quantitative course evaluations may provide more persuasive support for that feeling.

In our department, student evaluations of composition instruction involve a fifteen-question survey that employs a five-point scale. Each of us scored somewhere between 4.0 and 5.0 on all fifteen questions for both classes (Groups A and B). Collectively, we surpassed the departmental mean a total of 56 times, slightly more than half of a possible 90 (3 instructors x 2 classes x 15 questions). Thus, one could say that we scored well in a department in which students generally give high evaluations to composition instructors.

The remarkable fact is that, despite the similarity in the two groups' semester averages (2.63 vs. 2.68), all three of us received higher evaluations from Group B students. In other

words, regardless of what their final grades may indicate, students in Group B felt more strongly that they had benefited from our instruction. And although their evaluations covered the entire course, it is important to note that the survey was administered the week that we visited each other's classes.

To be specific, Group B students rated one of us higher than their Group A counterparts 34 times--more than three fourths of the 45 total comparisons (3 instructors x paired responses to 15 questions). Group B students scored all three of us higher in regard to six questions: Were the goals of the course clearly explained? Did the instructor's assignments lfill the goals of the course? Was prewriting helpful? Was revi ng helpful? Were the instructor's comments on papers helpful? Was the basis for grades clearly explained? The last question is, of course, the one that interests us most. Here, the differences ranged from relatively insignificant (4.50 vs. 4.42), to moderate (4.54 vs. 4.35), to the better part of a standard deviation of .67 (4.82 vs. 4.41).

The first two questions are also interesting, since they bear on the function of English 101 within the university curriculum and thus, tangentially, on the student's self-confidence in regard to academic discourse. Question 2 (Did the instructor's assignments fulfill the goals of the course?) yielded the clearer contrast: 4.32 vs. 4.22; 4.62 vs. 4.41; and 4.39 vs. 4.06, the last more than half a standard deviation. Finally, questions 3 (Was prewriting helpful?) and 4 (Was revising helpful?) clearly bore on the goals of our project, but

differences in mean scores were small.

Though not under any illusion that we have presented definitive proof of anything, we believe that our findings provide a warrant for further inquiry. While compositionists have grown skeptical of empirical research models, those models still hold sway in regard to assessment. For instance, the 1991 edition of The Bedford Bibliography for Teachers of Writing lists fifteen titles under the heading "Response and Evaluation," of which eight are concerned specifically with assessment. Of those eight, six have what might accurately be termed an empirical orientation. Such terms as reliability, validity, syntactic complexity, and competency testing abound. Of the two remaining titles, Belanoff and Elbow's article on portfolios is fairly restricted in scope, leaving White's Teaching and Assessing Writing, published seven years ago, as the only title whose annotation explicitly connects assessment with social context. Compare this to the fifteen titles listed under "Composing Processes," an area surely no less influenced by empirical research during the late seventies and early eighties. No fewer than seven annotations contain such terms as social, context, ethnographic, case study, discourse community, race, gender, class, and culture.

We certainly do not wish to argue that the contributions of empirical research to the field of assessment have come to little and must therefore be supplanted by a whole new body of work. (If nothing else, our references to standard deviations a few paragraphs ago would be an odd contradiction were that the case.)

Rather, we would like to see ethnographic studies of assessment conducted with a comparable degree of rigor, complexity, and elegance.

In the meantime, the pedagogical implications of our study seem clear. Evaluation can be demystified when cross-grading partners define themselves as an interpretive community. And when, under such a circumstance, they discover that students, like the perceptive malcontent quoted earlier, constitute fundamentally different interpretive communities, cross-graders can demonstrate their reading strategies and acknowledge their critical biases. The ensuing dialogue should be enriching for both students and their instructors.

Works Cited

Belanoff, Pat, and Peter Elbow. "Using Portfolios to Increase Collaboration and Community in a Writing Program." WPA 9.3 (1986): 27-40.

Cooper, Charles R. "Holistic Evaluation of Writing." Evaluating Writing: Describing, Measuring, Judging. Ed. Charles R. Cooper and Lee Odell. Urbana, IL: NCTE, 1977. 3-31.

Knoblauch, C. H., and Lil Brannon. Rhetorical Traditions and the Teaching of Writing. Upper Montclair, NJ: Boynton, 1984.

White, Edward M. Teaching and Assessing Writing. San Francisco: Jossey, 1985.